UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

IN RE: PHARMACEUTICAL INDUSTRY AVERAGE WHOLESALE PRICE LITIGATION	)))))	MDL NO. 1456 CIVIL ACTION: 01-CV-12257-PBS Subcategory Docket: 06-CV-11337-PBS
THIS DOCUMENT RELATES TO	))	Judge Patti B. Saris
U.S. ex rel. Ven-A-Care of the Florida Keys, Inc. v. Abbott Laboratories, Inc., et al., No. 06-CV-11337-PBS	))))	Magistrate Judge Marianne B. Bowler

ABBOTT LABORATORIES INC.'S NOTICE OF FILING IN RESPONSE TO THE CONFIDENCE INTERVALS SUBMITTED BY DR. MARK G. DUGGAN

On February 5, 2010, the United States filed its Notice of Filing of Confidence Interval Calculations (Dkt. No. 6899). The United States' notice attached a letter from Dr. Mark G. Duggan containing "confidence interval" calculations in the above-captioned matter. Dr. Duggan's submission came in response to the Court's inquiry during the January 22, 2010 *Daubert* hearing. Prior to this response to the Court's concern that he did not provide confidence intervals for his extrapolations, Dr. Duggan had not provided confidence intervals for any of his extrapolations.[1]

For the reasons articulated in the attached Response of Dr. James W. Hughes to the Abbott Confidence Interval Calculations Submitted by Dr. Mark G. Duggan (and in many statistics texts referenced in Dr. Hughes's response), the "confidence intervals" computed by Dr. Duggan are statistically invalid and of no scientific validity. In addition to the statistic texts referenced by Dr. Hughes, Abbott respectfully directs the Court's attention to the decisions in

[1] Dr. Duggan's letter does not contain any confidence intervals for his Medicare extrapolations. Notably, on the same day that the United States submitted Dr. Duggan's "confidence intervals" for Medicaid, the United States abandoned Dr. Duggan's Medicare extrapolations. (*See* Dkt. No. 6897.)

Univ. Computing Co. v. Mgmt. Sci. Am., Inc., 810 F.2d 1395 (5th Cir. 1987) and *Chavez v. IBP, Inc.*, No. CV-01-5093-RHW, 2004 WL 5520002 (E.D. Wash. Dec. 08, 2004).

In *University Computing*, the court stated:

> 'Tests of statistical significance and confidence intervals are mathematical procedures which yield probability statements relating the numerical characteristic of a sample population with the characteristics of a larger population or "universe" from which the sample was drawn. Tests of statistical significance and confidence intervals *always* presuppose a sample and a universe from which the sample is drawn. Consequently, the strict interpretation of results produced by either method requires one to consider the applicability of that presupposition to the case at bar.' (emphasis in original).

810 F.2d at 1399 (quoting D. Baldus & J. Cole, *Statistical Proof of Discrimination* 305 (1980)).

In *Chavez*, the court rejected an expert's "multiplier technique" based on a non-random, non-probability sample, in part because he could not provide a valid confidence interval. The court reasoned:

> Defense expert Dr. Javitz, a statistician . . ., worked to cast doubt on [plaintiff expert] Dr. Mericle's sampling methodology. At trial, Dr. Javitz opined that probability sampling is the only scientifically valid method to quantify the reliability of a sample because it is the only method that allows a statistician to calculate a confidence interval. . . . By failing to engage in probability sampling, Dr. Javitz explained that Dr. Mericle could not know his study's rate of error. Thus, any attempt to extrapolate the results from his 27 subjects to the entire Pasco Plant could have an unacceptably high rate of error, or contain unquantifiable biases.
>
> * * *
>
> The Court finds that the generally-accepted methodology in statistics and survey research is to conduct probability sampling. The Court remains unconvinced that Dr. Mericle's methodology of approaching employees every five minutes sufficiently randomized the sample. Instead, this method increased the study's unreliability and introduced new factors of bias. *Part of the reason for conducting probability sampling is to enable the surveyor to determine the study's rate of error*. See Daubert, 509 U.S. 579 at 590. Dr. Mericle did not engage in random sampling, and this undermines the reliability of the Chavez report.

Chavez, 2004 WL 5520002 at *9 (emphasis added).

Dated: February 24, 2010

Respectfully submitted,

/s/ James R. Daly

Daniel E. Reidy
James R. Daly
Jason G. Winchester
Brian J. Murray
JONES DAY
77 West Wacker Drive, Suite 3500
Chicago, Illinois 60601
Telephone: (312) 782-3939
Facsimile: (312) 782-8585

R. Christopher Cook
David S. Torborg
JONES DAY
51 Louisiana Avenue, N.W.
Washington, D.C. 20001-2113
Telephone: (202) 879-3939
Facsimile: (202) 626-1700

Counsel for Defendant Abbott Laboratories Inc.

CERTIFICATE OF SERVICE

I, David S. Torborg, an attorney, hereby certify that I caused a true and correct copy of the foregoing to be served on all counsel of record electronically by causing same to be posted via LexisNexis, this 24th day of February, 2010.

/s/ David S. Torborg
David S. Torborg

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

IN RE PHARMACEUTICAL INDUSTRY AVERAGE WHOLESALE PRICE LITIGATION	)	MDL NO. 1456 Civil Action No. 1-12257-PBS
THIS DOCUMENT RELATES TO: *United State of America ex rel. Ven-a-Care of the Florida Keys, Inc., v. Abbott Laboratories, Inc.*, No. 06-CV-11337-PBS	)	MASTER FILE NO. Judge Patti B. Saris

RESPONSE OF DR. JAMES W. HUGHES TO THE ABBOTT CONFIDENCE INTERVAL CALCULATIONS OF DR. MARK G. DUGGAN

SUMMARY

1. On February 5, 2010, the United States filed a Notice of Filing of Confidence Interval Calculations. That filing submitted a letter from Dr. Duggan (the "Duggan Letter") in which he presents "confidence intervals" around his estimate of extrapolated Medicaid damages presented in his reports of June 19, 2008 and January 23, 2009. This submission seeks to explain to the Court why Dr. Duggan's "confidence intervals" are statistically unsound and incapable of validating Dr. Duggan's 38-state extrapolation.

- First, as discussed during the *Daubert* hearings but revisited below to address the Court's concerns, it is uniformly recognized in the statistics literature that the type of confidence interval Dr. Duggan computes *requires* the use of a random probability sample, which Dr. Duggan admittedly did not use in his Medicaid extrapolation. The lack of a random probability sample is not ameliorated by the number of claims that he analyzed, or by his subsequent analysis of 9 additional states.

- Second, Dr. Duggan uses incorrect boundaries for his confidence intervals. When a confidence interval is based on an estimate of standard deviation, as is the case here, the boundaries of the confidence interval depend on the sample size. Because Dr.

Duggan's extrapolations rely upon average "difference fractions" he computed from a sample of 10 or fewer states, he cannot utilize the boundaries of the normal distribution (1.65 times for a 90% confidence interval, 1.96 times for a 95% confidence interval) in computing his confidence intervals. Because of his small sample size, the correct critical values will be higher than those used by Dr. Duggan to achieve a particular level of confidence. Using the higher critical value figures would serve to widen Dr. Duggan's confidence intervals significantly.

- Finally, Dr. Duggan improperly reduces his calculated variance by ignoring sources of variation in his estimates that he acknowledges exist.

BECAUSE DR. DUGGAN DID NOT USE A RANDOM PROBABILITY SAMPLE, HIS CONFIDENCE INTERVALS ARE STATISTICALLY INVALID AND UNRELIABLE

Application of Fundamental Sampling Principles to Dr. Duggan's Methodology

2. Sampling is the process of selecting a part of the population for the purpose of gaining information about the entire *population*. The *population* is "[a]ll of the units of interest to the researcher."[1] A sample is drawn from a *sampling frame*, a "list of units designed to represent the entire population as completely as possible."[2] In order to make valid inferences about the population based on any sample, the researcher must assure that the sampling frame actually represents the population of interest.[3] Researchers usually assure such an adequate connection between the sampling frame and the population of interest by using some type of a "random," or "probability," sample. "Probability samples have the distinguishing characteristic that each unit in the population has a known, nonzero probability of being included in the sample."[4]

---

[1] Donald H. Kaye and David A. Freedman, Reference Guide on Statistics (2d ed.) at 168 (definition of "population").
[2] Id. at 172 (definition of "sampling frame").
[3] Id. (sampling frame should "represent the entire population as completely as possible").
[4] Gary T. Henry, Practical Sampling, (Sage Publications) 1990 at 25; see also Michael O. Finkelstein and Bruce Levin, Statistics for Lawyers, 2nd ed. (New York, Springer) 2001 at 256.

3. Probability samples are different from nonprobability samples. Nonprobability sampling is any sampling method where some elements of the population have no chance of selection, or where the probability of selection cannot be accurately determined.[5] One type of nonprobability sample is commonly referred to as a "convenience" or "haphazard" sample, which is selected based on what is readily available and convenient.[6] The use of nonprobability sampling, such as a convenience sample, places limits on how much information a sample can provide about the population, making it difficult to extrapolate from the sample to the population.[7] Thus, because the nonrandom sample "will likely be biased in unknown ways . . . inferences from nonprobability samples are very risky."[8]

4. As established at the *Daubert* hearing, Dr. Duggan admittedly used a nonrandom convenience sample in his extrapolations. See 1/22/10 hearing at 214 ("I will be the first to admit it's not a random [sample]").

5. Dr. Duggan's across-state Medicaid extrapolations proceeded as follows.

- *First*, he used a nonrandom set of detailed claims data from 10 high-expenditure states to perform a "claim-by-claim" difference analysis for those states. Due to the unavailability of data from most of these states, his claim-by-claim analysis was limited to varying portions of the 1991-2001 damage period.[9] Despite Dr. Duggan's

[5] David Freedman, Robert Pisani, Roger Purves, Statistics, (W.W. Norton Co., New York) 1978 at 366.
[6] Finklestein and Levin, supra. note 4 at 257.
[7] Henry, supra. note 4 at 24 ("Because of the subjective nature of the selection process, nonprobability samples add uncertainty when a sample is used to represent the population as a whole. Confounding variables can influence the study results.")
[8] Bruce A. Thayer, The Handbook of Social Work Research Methods, (Sage Publications, 2001) at 49.
[9] Attached as Exhibit 1 is a chart showing the states and time periods included in Dr. Duggan's 10-state Medicaid sample.

characterization of his sample as a "10-state sample," he had data on all ten states for only 5 of the 44 quarters analyzed.

- *Second*, for each of these 10 states, Dr. Duggan computed, on a NDC-quarter specific basis, a "difference fraction." The difference fractions represent the percentage by which a state's spending would purportedly decrease had his revised prices been used.
- *Third*, Dr. Duggan computed average difference fractions, specific to each NDC and quarter at issue, by taking a simple average of those state-specific difference fractions he computed. Because he did not have complete data for the 10 states covering the entire 1991-2001 time period, Dr. Duggan's average difference fractions are usually computed from less than 10 state-specific difference fractions. Early in the period his extrapolations were based on considerably fewer states; for example, his extrapolations for 1991 were based on only one state.
- *Fourth*, Dr. Duggan computed NDC-quarter-specific "differences" for each of the remaining 38 states by multiplying those states' NDC-quarter-specific expenditures, obtained from CMS data (SMRF/MAX or SDUD), by his NDC-quarter-specific average difference fractions.[10]

6. In the "sample" that Dr. Duggan used in his across-state extrapolation, no claims submitted to the vast majority of states had any chance of being included in Dr. Duggan's sample. The type of sample that Dr. Duggan employed for his Medicaid extrapolation is

[10] Indiana presents a unique case, as Dr. Duggan extrapolated a difference for that state using only the difference fractions from Illinois.

commonly referred to in statistics texts as a "convenience" of "haphazard" sample.[11] Worse, here there is a complete disconnect between Dr. Duggan's *sampling frame* (claims from 10 or fewer states) and the purpose of his sampling exercise—to estimate differences in 38 different states.[12]

7. An example may help illustrate the fundamental problem in Dr. Duggan's methodology. Assume a researcher wants to use sampling to estimate the percentage of defects experienced by United States customers during the first three years of owning high definition televisions (HDTVs) made by each of 9 Japanese manufacturers (Sony, Panasonic, Toshiba, Sharp, JVC, Pioneer, Hitachi, Mitsubishi, and Sanyo) from 1995 to 2005. The *population* of interest is all Japanese HDTVs sold in the United States from 1995 to 2005. One general approach, consistent with the principle of probability, would be to (1) identify all United States customers who purchased Japanese HDTVs during the relevant time period as the sampling frame, (2) assign each a control number, (3) use random selection to identify a subset of customers to be included in the sample, and (4) survey those customers regarding any defects experienced.[13]

---

[11] Finklestein and Levin, supra. note 4 at 257 ("Note that 'haphazard' or 'convenience' samples are *not* random samples.").

[12] In preparing his report, Dr. Duggan did very little to establish that, for purposes of his calculation, the sampled states were representative of the unsampled states. Most of his observations, for example noting the similarity of the basic adjudication formulas, have been by explicitly rejected by the OIG as being indicative of the comparability of states' prices for generic drugs. See Variation in State Medicaid Drug Prices (OEI-05-02-00681, 9/2004) (discussed during 1/22/10 hearing at 131-38). The only thing of substance—his comparison of "per-claim" spending across the states (which covered only the 1999-2001 time period)—was relegated to *a footnote* in his initial report (footnote 45) and was thoroughly discredited as a measure of representativeness across the states at the hearing. See 1/22/10 hearing at 138-48; 12/11/10 hearing at 75-80.

[13] Of course, more sophisticated sampling methods—such as stratified or cluster sampling—might be used to better assure that the sample contains a representative mix of models, years, and manufacturers.

8. A different approach, inconsistent with the principle of probability (but similar to the approach employed by Dr. Duggan), would be to survey each United States customer who purchased HDTVs manufactured by three high-volume Japanese manufacturers (say, Sony, Panasonic, and Toshiba). One might reason (similar to Dr. Duggan) that these manufacturers collectively represent a majority of Japanese HDTVs sold in the United States, and, after all, a larger sample size must be better. While this approach would provide a valid estimate of the number of defects experienced *by owners of Sony, Panasonic, and Toshiba HDTVs* (as each and every purchaser of these brands was surveyed)—and these brands do represent a significant *part* of the total population of interest—this approach *cannot* under any reasoning provide a valid estimate of the number of defects experienced by owners of Sharp, JVC, Pioneer, Hitachi, Mitsubishi, and Sanyo HDTVs or even of the population as a whole. This approach is effectively the methodology that Dr. Duggan employed in his Medicaid extrapolation. For similar reasons, Dr. Duggan's approach does not result in a statistically valid estimate of the impact of his revised prices in states where he did not utilize any detailed claims information.[14]

9. In sum, the approach that Dr. Duggan employed in his report to compute "differences" for the 38 states really does not resemble any sort of sampling at all. There is no discussion of sampling procedures. There is no discussion of standard deviations, standard errors, confidence intervals, or other benchmarks commonly associated with sampling. Dr. Duggan makes no attempt to weight or otherwise adjust his nonrandom sample to better resemble the population of interest. Rather, it appears that Dr. Duggan and

[14] The CMS data (SMRF/MAX and SDUD) does not contain the information necessary to compute a "difference" even under Dr. Duggan's methodology. Tellingly, Dr. Duggan himself did not use that information to compute his difference for the 38 states. He instead used the findings from the ten states.

Plaintiffs focused on computing differences for a select number of high-expenditure states. A full 46 pages of the text of Dr. Duggan's report is dedicated to the "differences" he computed for the 10 (then 11, as it included Ohio) sample states, while his 38-state extrapolation is explained in a scant 4 pages. The lack any discussion of sampling procedures or confidence intervals suggests that Dr. Duggan's 38-state extrapolation was an afterthought, devoid of meaningful sampling analysis.

Confidence Intervals Require the Use of Probability Samples

10. When one uses statistical methods to make inferences about a population based on a sample (which is what Dr. Duggan does in his Medicaid extrapolations), it is standard practice to calculate a confidence interval as a measure of the accuracy of a statistical point estimate.[15] As described in the Elementary Statistics text, a "confidence interval is a range of values that is *likely* to contain the true value of the population parameter."[16] Stated another way, a "level *C* [say, 95%] confidence interval for a parameter is an interval computed from the sample data by a method that has a probability *C* [95%] of producing an interval containing the true value of the parameter."[17] The greater the confidence level desired, the wider (or larger) the confidence interval will be.

11. Even when a statistically valid sample is used, the sample mean will vary from the population mean.[18] This is known as "random error," "sampling error," or "sampling

---

[15] See Mario Triola, Elementary Statistics, 7th ed. (Reading, MA, Addison-Wesley) 1998 at 289-90.
[16] Id.
[17] David S. Moore and George P. McCabe, Introduction to the Practice of Statistics, 4th ed. (New York, W.H. Freeman & Co.) 2002 at 420.
[18] Reference Guide, supra. note 1 at 116 ("More generally, a sample is unlikely to be a perfect microcosm of the population; statisticians call differences between the sample and the population, just due to the luck of the draw in choosing the sample, "random error" or "sampling error.").

variation," and is minimized by the use of valid sampling procedures. As explained in the *Reference Guide to Statistics*, a confidence interval *quantifies only random sampling error*.[19] The specific formulas developed over the decades for computing confidence intervals are readily found in each and every elementary statistics textbook. There is no debate about what the formulas are, how they are to be used, and their limitations. As stated in the *Reference Guide,* standard errors and confidence intervals simply do "not address problems inherent in using convenience samples rather than random samples."[20]

12. For this reason, authoritative statistics texts make clear that statistical inferences—including the use of confidence intervals—from convenience samples are simply inappropriate. For example:

> **David Freedman, Robert Pasani & Roger Purves, Statistics, (New York, W.W. Freeman) 1978 at 350, 355, 366**
>
> > "Warning. The formulas for simple random samples should not be used for other kinds of samples."
> >
> > "When a sample is taken by a probability method it is possible not only to estimate the parameter, but also to figure the likely size of the chance error in the estimate."
> >
> > The formulas for simple random samples shouldn't be used for other kinds of samples. *If the sample was not chosen by a probability method, there is no sensible way to calculate standard errors for it.*" [emphasis added].
> >
> > A *sample of convenience* is a sample that isn't chosen by a probability method. . . . Some people use the simple random sample formulas on

---

[19] Id. at 119-21 ("Standard errors and confidence intervals are derived using statistical models of the process that generated the data. . . . Our exampled was based on a random sample, and that justified the statistical calculations. . . . Standard errors and confidence intervals generally ignore systematic errors such as selection bias or non-response bias; in other words, these biases are assumed to be negligible. . . . [T]he standard error does not address problems inherent in using convenience samples rather than random samples.").
[20] Id. at 121.

> samples of convenience. *That is a real blunder. With samples of convenience, there is no way to define the chances, and therefore no way to get standard errors*." [emphasis added]

Bruce A. Thyer, <u>The Handbook of Social Work Research Methods</u>, (Sage Publications, 2001) at 42, 49

> "Probability samples have at least two advantages. First, they are unbiased and representative of the population. *Second, we can estimate the amount of error involved in using the statistics we get from the samples as estimates of the values we would obtain if we observed the entire population*." [emphasis added]

> "Probability samples can be contrasted with nonprobability samples. When using nonprobability sampling methods, we cannot estimate the probabilities associated with selection of different samples of size *n*. This means that any of a plethora of biasing factors may be operative . . . The major consequences of this are that (a) the sample will likely be biased in unknown ways and *(b) we cannot estimate the error involved in our use of sample statistics as estimates of population parameters. Thus inferences from nonprobability samples are very risky*." [emphasis added]

David S. Moore and George P. McCabe, <u>Introduction to the Practice of Statistics</u>, 4th ed. (New York, W.H. Freeman & Co.) 2002 at 416

> "*When you use statistical inference you are acting as if the data come from a random sample or a randomized experiment.* If this is not true, your conclusions may be open to challenge. Do not be overly impressed by the complex details of formal inference. This elaborate machinery cannot remedy basic flaws in producing the data. . . . " [emphasis in the original]

David S. Moore and George P. McCabe, <u>Introduction to the Practice of Statistics</u>, 4th ed. (New York, W.H. Freeman & Co.) 2002 at 426

> "There is no correct method for inference from data haphazardly collected with bias of unknown size. Fancy formulas cannot rescue badly produced data."

Gary T. Henry, <u>Practical Sampling</u>, (Sage Publications, 1990) at 24

> Because of the subjective nature of the selection process, nonprobability samples add uncertainty when a sample is used to represent the population as a whole. Confounding variables can influence the study results. *The accuracy and precision of statements about the population can only be determined by subjective judgment*." [emphasis added]

Michael O. Finkelstein and Bruce Levin, Statistics for Lawyers, 2nd ed. (New York, Springer) 2001 at 256-7:

> "The statistician's reason for preferring a random sample . . . is to be able to make probabilistic statements, such as the following: 'the estimated proportion of improper allowances in the sampled population is 0.10, and with 99% confidence the proportion is at least 0.05, in the sense that if the true (population) proportion were any less, there would be less than a 1% chance of observing a sample with as high a proportion of improper allowances as was in fact observed.' *Without the probability space generated by the random sampling procedure it becomes impossible, or at best speculative, to assign quantitative statements about the reliability of inferences based on the sample*." [emphasis added].

William Medenhall, Robert Beaver & Barbara Beaver, Introduction to Probability and Statistics (Brooks/Cole 13th ed 2009) at 258

> "Remember that nonrandom samples can be described but cannot be used for making inferences!"

13. Several of these texts are cited in a useful article by Michael C. Mosier, a Professor in the Department of Mathematics and Statistics at Washburn University. In that article, A Real Data Approach to Teaching the Consequences of Using A Nonrandom Sample (attached as Exhibit 2), Professor Mosier explains:

> "When teaching statistics to students who are not majoring in the subject, we are often constrained to limit the amount of theoretical treatment certain topics receive. Most of the information provided to students concerning random selection of samples is focused on methods of obtaining random samples. We don't have the time to develop the theoretical reasons why it is so important. Unfortunately, when students leave our introductory courses and go on to perform research in other fields, they are often satisfied with collecting and analyzing data obtained from non-random samples. It is likely they will not realize that statistical inference performed on such data cannot be trusted. For example, it is all too common for beginning researchers to construct a 95% confidence interval for a mean, using data collected from a convenience sample, and have no idea that the actual confidence level of the interval is quite likely far below 95%.

> Most introductory statistics textbooks make the point clearly, that random sampling is one of the conditions that must be met in order for inferential procedures to be valid . . . . *All texts agree that random sampling is an important condition that must be met for the capture rate of confidence intervals to be correct* [emphasis added]."

Despite my testimony during the Daubert Hearing consistent with these statistics texts, Dr. Duggan offers no contrary evidence from learned treatises or any other source that his use of a convenience sample is of no consequence to the accuracy and validity of his confidence interval calculations. In my opinion, no such impartial evidence exists.

14. Because Dr. Duggan used a convenience sample, rather than a random probability sample, his confidence intervals are on that basis alone statistically invalid and not consonant with accepted statistical practice.

15. A confidence interval is mathematically driven by the *variance* in the underlying population and the sample size.[21] The greater the variance in the underlying population, the wider (or larger) the confidence interval will be (meaning the lower bound will be a lower figure).[22] As discussed during the Daubert Hearing, the major problem with convenience sampling, as Dr. Duggan uses for all of his calculations, is that it introduces "nonsampling variability" of *unknown and unquantifiable size*. As described in the Statistics for Lawyers text, nonsampling variability includes all of the variation introduced by defects in the selection of the sample:

> "In practice, sampling variation is rarely the only source of variability. The other principal sources are (i) defects in the sampling frame involving incomplete or inaccurate enumeration of the target populations; (ii) defects in the methods of selection that result in unequal probabilities of selection; and (iii) defects in the collection of data from the sample (including such matters as nonresponse bias, evasive answer, and recall bias).

[21] Finklestein and Levin, supra. note 4 at 261.
[22] Id.

> A common misconception is that a sample's precision depends on its size relative to the size of the population. . . . The more correct view is that if a population can be sampled correctly, the precision of the sample depends principally on the variability of the population, to a lesser extent on the sample size, and perhaps most importantly on the avoidance of defects that create nonsampling variation."[23]

Similarly, David DeVaus, in Analyzing Social Science Data, writes:

> "Sample bias is always a potential problem in any sample-based research. Bias occurs when the characteristics of a sample are systematically different from those of the population. . . . Bias is a problem because it:
>
> - introduces error when population estimates are being made from the sample;
> - can distort relationships between variables within a sample.
>
> Sample bias arises for one or more of the following reasons:
>
> - selecting a study population that is not an adequate reflection of the theoretical population;
> - having a biased sampling frame;
> - biased non-response patterns . . ."[24]

16. The convenience sample that Dr. Duggan employs in his Medicaid calculations includes each of the defects described above. Dr. Duggan's sample omits claims from 38 states entirely, and also does not include any claims from significant portions of time for most of the 10 states included in his sample. Such claims had no probability of being included in his sample.

17. As a practical matter, if we had reason to believe that the impact of Dr. Duggan's revised prices would be consistent across the 50 states (i.e., his "difference fractions" would all be roughly the same), the concerns raised in the above texts would not be as troublesome. However, on the relevant question (the impact of Dr. Duggan's revised prices), we have

---

[23] Finkelstein and Levin, supra. note 4 at 261.
[24] David DeVaus, Analyzing Social Science Data, (London, Sage Publications) 2002 at 152.

many reasons to believe that the population of states is not homogenous. As summarized in a 2004 OIG report, it is well known that states' payments for generic drugs vary considerably—even when states use the same AWP-based formula.[25] As discussed at length during the hearing, and as admitted in Dr. Duggan's recent confidence interval submission, Dr. Duggan's own difference calculations across the 19 states for which he has now performed some "claim-by-claim" analysis show significant variability across the states in the impact of his revised prices.[26]

18. In sum, because the confidence interval depends in large part on the variability of the entire population of interest, and because Dr. Duggan does not know and could not calculate that variability for the non-sampled states and time periods, his confidence intervals are invalid. In short, without any idea of the degree and direction of nonsampling variation in Dr. Duggan's convenience sample—and he provides none—his extrapolations and his calculation of a confidence interval around his extrapolations are of unknown accuracy and reliability.

Neither the Number of Claims That Dr. Duggan's Reviewed Nor His New Nine-State Analysis Validate His Extrapolations or Confidence Intervals

19. Any assertion that the problem of using a convenience sample is irrelevant, or remedied by the large number of claims he reviewed, is soundly rejected in even the most elementary statistics textbooks. [27] For example:

---

[25] See supra. note 12.

[26] See 1/22/10 hearing at 71-78; Duggan Letter at ¶14 (noting that "eight of the state effect coefficients are statistically significant at the 95% level, indicating that there is systematic variation across states").

[27] I note that Plaintiffs' counsel's attempt during the Daubert hearing to excuse Dr. Duggan's use of a non-random sample by invoking the Law of Large Numbers (i.e., as sample size increases, the sample mean approaches the population mean) is unavailing, as the Law of

Mario Triola, Elementary Statistics, 11th ed. (Boston, MA, Addison-Wesley) 2010 at 26

> "If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical torturing can salvage them"

Mario Triola, Elementary Statistics, 7th ed. (Reading, MA, Addison-Wesley) 1998 at 288

> "Data collected carelessly can be absolutely worthless, even if the sample is quite large."

The fact that Dr. Duggan used very complete data for certain time periods in certain states that accounted for a large percentage of the total number of claims at issue simply reflects that Plaintiffs and Dr. Duggan were able to obtain certain data about a *different population* (originally 10 states; a total of 19 states after Dr. Duggan's subsequent 9-state analysis). It does *not* tell us that those states and time periods are representative of the other states and time periods, or allow us to draw valid statistical inferences or compute valid confidence intervals regarding the remaining states. This is particularly true because of the extensive non-sampling variability that exists in Dr. Duggan's sample of convenience.

20. In paragraph 17 of his confidence interval submission, Dr. Duggan discusses an analysis of 9 additional nonrandomly selected states that he performed after Abbott lodged its Daubert challenge. He observes that his claim-by-claim analysis computed a difference of

---

Large Numbers applies only to random samples. The definition of the Law from an elementary statistics text states:

> "Draw independent observations *at random* from any population with finite mean μ. Decide how accurately you would like to estimate μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close" [emphasis added].

Moore and McCabe, supra. note 17 at 322.

$4.030 million, versus his extrapolation of $4.106 million. While Dr. Duggan's subsequent analysis may provide some level of assurance regarding the reliability of his difference calculation for certain time periods for those 9 additional states (if you ignore other underlying flaws in his overall difference approach), it provides no assurance whatsoever regarding the reliability of his extrapolations for the remaining 29 states. In fact, Dr. Duggan's new analysis showed, for individual states, that his extrapolations could be overstated by more than 40%.[28] When asked if he knew whether the remaining 29 states were similarly overstated, Dr. Duggan admitted he did not know.[29] In sum, neither Dr. Duggan's confidence intervals, nor his subsequent nine state analysis provides any valid, quantitative level of confidence for the reliability of his estimates for the 29 remaining states.[30]

21. Here again it is instructive to return to my example of using sampling to estimate the percentage of defects experienced on HDTVs manufactured by 9 Japanese television makers. Dr. Duggan's new 9-state analysis is analogous to the researcher in our example checking the reliability of the original estimates for 2 manufacturers not included in the sampling frame (say, Sharp and JVC) by performing a subsequent survey of all of those manufacturers' customers. Suppose these subsequent surveys showed that the researcher's original defect estimates for Sharp were understated by 30%, while the estimates for JVC were overstated by 30%. Although the subsequent surveys provide information on 2 more manufacturers—

---

[28] 1/22/10 hearing at 70, 117, 149-50, 197.
[29] 12/11/09 hearing at 72-73.
[30] It should also be noted that neither Dr. Duggan's 9-state analysis nor his confidence intervals have any application to the extrapolated difference he calculated for Indiana. For that state, Dr. Duggan extrapolated using only the results of his claim-by-claim analysis of Illinois. Dr. Duggan has not computed any sort of confidence interval pertinent to the reliability of his Indiana extrapolation.

and happen to roughly even out in the end—the subsequent surveys tell us nothing whatsoever, statistical or otherwise, regarding the reliability of the researcher's estimate of defects for the 4 remaining manufacturers (Pioneer, Hitachi, Mitsubishi, and Sanyo). Like Dr. Duggan, our researcher does not know if the original estimates were fairly close, wildly inaccurate, or spot on.

The Haphazard Nature of Dr. Duggan's Confidence Interval Calculations Highlights the Flaws in His Approach

22. Both Dr. Duggan's approach to computing his confidence intervals, as well as his subsequent 9-state analysis, further highlight the fundamental problem of basing a confidence interval on a haphazard convenience sample.[31] For example, I note that Dr. Duggan includes in his calculation of the variance "difference fractions" for certain NDC-quarters that *use only one state*, Illinois. A full 9% of Dr. Duggan's "difference fractions" is based *solely on Illinois*. The variance in this "average" is exactly zero, as the "average" is calculated with only one observation. Clearly, estimates calculated from only one state do not have zero variation—if this were valid, all samples would include only one observation. By computing a zero variance for this one state "average," Dr. Duggan effectively assigns a zero confidence interval for these NDC-quarter difference fractions, improperly suggesting he is 100% confident in the point estimate of differences he computed for those quarters. Dr. Duggan's inclusion of these aberrant observations serves to artificially lower his variance estimate and illegitimately narrow his confidence interval calculation.

23. Dr. Duggan's subsequent 9-state analysis demonstrates the fallacy of effectively assigning a zero confidence interval for this time period. His new analysis contains one state,

[31] It should be noted that Dr. Duggan's confidence intervals do not consider the results of his new 9-state analysis.

Connecticut, where he used detailed data for part of 1991. Across the 44 NDCs, his difference fractions for Connecticut are very different from his difference fractions for Illinois, suggesting a widely varying impact of his revised prices.[32] For example, on the highest expenditure NDC (00074653301), his Illinois difference fractions for Q2-Q4 1991 are 77.5%, 80.2%, and 81.1%, while the Connecticut rates for Q3 and Q4 1991 are 35.1% and 47.6%.[33] In other words, *had Dr. Duggan happened to have included Connecticut in his 10-state sample, his extrapolated difference rate for this NDC would have been about half of what it was, and his variance and confidence interval attributable to that NDC-year would have been much higher*.

24. Such anomalies exist throughout the claims period. As another example, for the third quarter of 1993, Dr. Duggan's average difference ratios and confidence intervals were based on 4 states (IL, NJ, NY & WI). For the second highest expenditure NDC (00074798302), his difference fractions were 79.4% (IL), 61.2% (NJ), 78.8% (NY), and 82.4% (WI).[34] However, Dr. Duggan's subsequent 9-state analysis included 3 more states with claims data for that time period, whose difference ratios were 94.9% (CT), 32.2% (IA), and 2.5% (MD).[35] Again, had Dr. Duggan happened to include these 3 additional states in his original "sample," his difference ratios and confidence intervals would be very, very different. Similarly, in the first quarter of 1997, Dr. Duggan's average difference ratios and confidence intervals were based on 8 states. For the third highest expenditure drug (00074650901), his

[32] Attached as Exhibit 3 is a spreadsheet showing the difference fractions that result from Dr. Duggan's "claim-by-claim" analysis for the 19 states for which he analyzed detailed claims data. This spreadsheet was contained at Tab 15 of the binder Abbott used during the January 22, 2010 hearing; it was sent to Plaintiffs on January 13, 2010.
[33] See attached Exhibit 3 at p. 6 of 36. Such stark differences are not limited to this NDC.
[34] Id. at p. 2 of 86.
[35] Id.

difference ratios were 77.8% (CA), 73.1% (FL), 72.9% (IL), 67.3% (KY), 76.1% (LA), 75% (NJ), 76.2% (NY), and 56.0% (WI).[36] Dr. Duggan's subsequent 9-state analysis included 4 more states with claims data for that time period, whose difference ratios were 49.4% (CT), 31.6% (MA), 32.2% (MD), and 71.9% (TX).[37] Once again, we can see that Dr. Duggan's average difference fractions and confidence intervals are dependent entirely on happenstance, not valid statistical methods. The foregoing illustrates the fundamental flaws of computing confidence intervals on such a haphazard convenience sample.

25. In sum, Dr. Duggan's extrapolation methodology was fundamentally flawed from the outset. And without the necessary foundation to compute a confidence interval—namely, a valid probability sample—Duggan's confidence intervals are equally flawed. In my opinion, Dr. Duggan's confidence intervals would never be accepted in any peer-reviewed publication for this reason alone.

DR. DUGGAN MISSTATES THE BOUNDARIES FOR HIS CONFIDENCE INTERVALS

26. Critically, Dr. Duggan uses the incorrect boundaries for his confidence intervals in paragraph 15 of his latest submission. In that paragraph, he states that the 90% confidence interval is "1.65σ" (i.e., 1.65 times the standard deviation) and that the 95% confidence interval is "1.96σ" (i.e., 1.96 times the standard deviation). This would be correct *if* the population were normally distributed and the standard deviation σ were known, which it is not here.[38] Dr. Duggan is clearly basing his confidence interval calculation on his *estimate* of the standard deviation, the calculation of which he describes in paragraphs 8-14 of his submission. When a confidence interval is based on an estimate of the standard deviation, as

[36] Id. at p. 17 of 36.
[37] Id.
[38] See Moore and McCabe, supra. note 17 at 492-94.

Dr. Duggan's is, the confidence interval is based on the *t* distribution, and the boundaries of the confidence interval are then based on the sample size.[39] The larger the sample size, the narrower the confidence interval, and vice versa.

27. Dr. Duggan's confidence interval calculation centers around his estimate of the variance of $\bar{d}_{nt}$, the average "difference fraction" across up to ten states included in his convenience sample. These are the difference fractions he applies to the 38 states in his extrapolations. Dr. Duggan indicates in equation (3) and ¶7 that the $\bar{d}_{nt}$ (the average "difference fractions") are calculated from a maximum of 10 individual states, and thus the variance for each $\bar{d}_{nt}$ (the average "difference fraction") is calculated from a maximum of 10 deviations from the mean, as shown in equation (4).

28. Thus, Dr. Duggan's confidence interval calculation must distinguish between those average "difference fractions" that are calculated based on one state, two states, up to ten states, as the boundaries of the confidence interval from the *t* distribution will vary according to sample size. Obviously, the confidence interval must be wider for "average" difference fractions calculated only on one state, Illinois, than it is for average difference fractions calculated on ten states. A wider confidence interval means that the lower bound of the confidence interval will be a lower figure.[40]

---

[39] Id.; Reference Guide, supra. note 1 at 118 n.117 ("Confidence levels are usually read off the normal curve (see infra Appendix). Technically, the area under the normal curve between –2 and +2 is closer to 95.4% than 95.0%; thus, statisticians often use ±1.96 SEs for a 95% confidence interval. . . . The normal curve gives good approximations when the sample size is reasonably large; for small samples, other techniques should be used."), 175 ("For small samples drawn at random from a population known to be normal, the t-statistic follows "Student's t-distribution" (when the null hypothesis holds) rather than the normal curve; larger values of t are required to achieve 'significance.'").

[40] The use of higher critical value figures (in place of 1.65 (90% confidence) and 1.96 (95% confidence)) does not ameliorate the failure to use a probability sample. In other words, one

29. This error is significant. Dr. Duggan calculated a total of 1775 average "difference fractions" by NDC, state and quarter. The table below shows the number and fraction of this total that were calculated based on one state, two states, up to ten states.[41] The table also shows the appropriate boundary for the confidence interval that Dr. Duggan should have used for those difference fractions calculated on a particular number of states.

# of States in DIFF_ FRAC	Number of DIFF_FRACs calculated using this number of states	Percent of DIFF_FRA Cs in this category	Critical value for 90% confidence interval (what Duggan should have used)	Multiplier Duggan used instead	Critical value for 95% confidence interval (what Duggan should have used)	Multiplier Duggan used instead
1	158	9%	6.314	1.65	12.71	1.96
2	166	9%	2.920	1.65	4.303	1.96
3	74	4%	2.353	1.65	3.182	1.96
4	109	6%	2.132	1.65	2.776	1.96
5	111	6%	2.015	1.65	2.571	1.96
6	139	8%	1.943	1.65	2.447	1.96
7	171	10%	1.895	1.65	3.365	1.96
8	375	21%	1.860	1.65	2.306	1.96
9	337	19%	1.833	1.65	2.262	1.96
10	135	8%	1.812	1.65	2.228	1.96

For example, for those average difference fractions calculated using only five states, the 90% confidence interval should have been calculated as 2.015 times the estimated standard deviation, rather than the 1.65 used by Dr. Duggan. Similarly, the 95% confidence interval for five state difference fractions should have been calculated as 2.571 times the estimated standard deviation, rather than the 1.96 used by Dr. Duggan. This correction would yield confidence intervals for these difference fractions significantly wider than those calculated by Dr. Duggan.[42]

---

cannot correct Dr. Duggan's confidence intervals by simply re-running his calculations using the higher critical value figures.

[41] Table excerpted from Moore and McCabe, supra. note 17, back cover.

[42] Dr. Duggan may claim that his sample size is 1775, rather than the one to ten states indicated here. Such a claim would be fallacious, insofar as, for example, the average

DR. DUGGAN'S CONFIDENCE INTERVALS ARE PREMISED ON INACCURATE ASSUMPTIONS

30. Finally, even if one were to ignore the foregoing defects, there are numerous other errors that invalidate Dr. Duggan's confidence interval calculations.[43] First, he makes key assumptions that he acknowledges are not accurate. In response to a question from the Court as to whether or not he could calculate confidence intervals for his DIFF_FRAC estimates, Dr. Duggan acknowledged that the DIFF_FRACs for the various drugs, states and quarters were correlated.[44] Yet in his calculation of the confidence interval, he assumes that the difference fractions are uncorrelated.[45] If, as he recognizes, these difference fractions are in fact correlated, then *the formula he uses to calculate the variance of the difference fractions (Equation 4 in his submission) is not the correct formula.* Where variables are correlated[46] the proper calculation of variance must take into account this correlation.[47] Dr. Duggan

---

difference fraction for any NDC for the second quarter of 1991 was calculated using only the difference fraction for the State of Illinois. Including a count of other states, quarters, or NDCs in the variance calculation would be an error, as no data from these other states, quarters or NDCs entered into the second quarter 1991 calculation.

[43] I note that even if his confidence interval calculation were correct, it would only account for sampling variability. Dr. Duggan's calculation does not account for any nonsampling variation.

[44] See 1/22/10 Daubert hearing at 207 ("But then I recognize that there is some correlation between these difference fractions . . . ."). .

[45] Duggan Letter at ¶8 ("If one assumes that these variables, d_{snt}, [the difference fractions] are uncorrelated . . . .").

[46] Correlated variables move together rather than independently. In the current setting, a high difference fraction for, say, saline in Illinois in 1998 implies a higher difference fraction for the other 43 NDCs for that state and year, as all the claims are adjudicated by the same agency using the same formula.

[47] When variables (e.g. difference fractions) are independent the variance of their sum is simply the sum of the variances. When variables are correlated, the variance of their sum is the sum of the variances, *plus two times the covariance of the variables*. Thus, as a practical matter, if these variables are positively correlated, correctly calculating the variance of these correlated variables will increase the variance and the confidence interval. See Moore and McCabe, <u>supra.</u> note 17 at 330.

ignores the correlation he acknowledges exists, thereby *reducing* his calculated variance of the "difference fractions," and improperly *narrowing* his confidence interval calculations.

31. In addition, Dr. Duggan has failed to adjust his confidence interval calculations for other sources of variation that he acknowledges exist. Had he properly done so, his confidence intervals would have been wider. In particular, in paragraph 8 of his most recent submission, Dr. Duggan states that "The value of d_{snt} [the difference fractions] varies across states, products and quarters . . ." In other words, the values of d_{snt} are not constant across any of these dimensions. Contrary to his own observation, Dr. Duggan assumes that the difference fractions are all drawn from the same distribution, and have the same standard deviation. Next, in paragraphs 9-14, he adjusts his variance calculation to allow for the possibility that the difference fraction for a particular drug and quarter may differ systematically by state. He finds that the difference fractions do indeed vary systematically across states[48], increasing his variance estimate, and widening his confidence interval. However, even after he rejects the assumption that the difference fractions all have the same standard deviation, he fails to make a similar adjustment in his variance calculation to allow the difference fractions to vary systematically across time or across NDCs. Given that he acknowledges that the difference fractions vary by quarters and products, making such needed adjustments would again increase his variance estimate and widen his confidence interval. Dr. Duggan offers no explanation as to why he chose to adjust for only one of three acknowledged sources of variation in his calculation.

---

[48] Duggan Letter at ¶14. Dr. Duggan estimates that the standard deviation of the "state effects" is 0.04, or 4%, but again, he offers no confidence interval on this estimate, so we have no idea where his estimate of the average deviation from the mean is 4% plus or minus 1% or plus or minus 20%.

CONCLUSION

32. Dr. Duggan's confidence interval calculations are completely invalidated by the fact that his methodology is only appropriate for random probability samples, which he admittedly does not use. Numerous authoritative statistics texts confirm this. Even if one overlooks the shortcomings of the sample, he constructs his confidence intervals using several baseless and unfounded assumptions that he himself knows to be false. As a result, his "confidence intervals" are neither statistically valid nor consonant with accepted statistical practice. In my opinion, Dr. Duggan's confidence intervals lie entirely outside accepted economic or statistical practice, and would not be accepted in any peer-reviewed publication.

February 24, 2010

Dr. James W. Hughes